

**REMARKS**

Applicants have studied the Office Action dated February 23, 2004. No new matter has been added. It is submitted that the application, is in condition for allowance. Applicants have amended claims 1 - 7, 10, and 14 - 20. By virtue of this amendment, claims 1-20 are pending. Reconsideration and further examination of the pending claims in view of the above amendments and the following remarks is respectfully requested. In the Office Action, the Examiner:

- Rejected claims 1-3, 10, 14-16, and 20 under 35 U.S.C. §103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702);
- Rejected claims 4-6, and 17-19 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Aganovic et al. (U.S. Patent Number 6,105,042);
- Rejected claim 7 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), in further view of Meyerzon et al. (U.S. Patent No. 6,199,081) and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314);
- Rejected claims 8-9 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Meyerzon et al. (U.S. Patent No. 6,199,081);
- Rejected claim 11 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314);
- Rejected claims 12 and 13 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), in further view of Meyerzon et al. (U.S. Patent No. 6,199,314), in further view of Hughes et al. (U.S. 5,892,908), and in further view of Aganovic (U.S. Patent No. 6,105,042).

ARC9-2000-0046-US1

9

09/607,370

### Overview of the Present Invention

The present invention provides a web crawling unit, a method, and a computer readable medium for assisting in the index of informational content at a given website by use of a search engine. One type of search program for indexing informational content at a website is known as a "web crawler." The web crawler creates an index of informational content for a given website for subsequent use by a search engine. Simple web pages where all the content is from one or two files the indexing by web crawlers is very useful. One problem often encountered by web crawlers is where information presented in a web page is stored in secondary pages. The use of secondary documents such as text, images, and other multimedia, make the management of web page content much easier because each component of a web page is broken down into pieces. Web crawlers do not load these secondary documents when indexing web page content and this informational content is not properly indexed for a search engine.

Another problem encountered with using web crawlers to index informational content, is where the web page is dynamically assembled. Dynamically assembled information is information presented on a single web page from more than one location. The dynamically assembled web content is not available from one location and must be gathered from several different storage locations before being presented as a single webpage to the user. This is especially true in website where part of the informational content is retrieved from a database. The contents of the database must be gathered and assembled into a webpage format. Further, the use of client side scripts such as JavaScript and VBScript, do not make all the information available until the script is executed on the client side. Since web crawlers index informational content on server sites as opposed to client sites, the information content in client side script is not captured. The present invention solves the problem of web crawling dynamic data documents by temporarily rendering the information for the dynamic data document in memory in the manner the composer of the document intended it to be displayed. In this manner, the subsequent crawling indexes all the information on the website in the manner in which the creator of the web site intended the site to be viewed. This is in

ARC9-2000-0046-US1

10

09/607,370

contrast to only discrete pieces of the informational content on the web site being indexed in the prior art. The present invention solves the problem with images containing textual data in part through use of optical character recognition on the images themselves to assist with web crawling.

To overcome the problems of using web crawlers to properly index dynamic websites containing secondary documents, with or without client-side scripts and the use of images with textual content, the present invention retrieves a web document at a given address or URL. The contents of the web document are extracted for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Next, secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation are loaded. The in-memory webpage representation is analyzed to produce a text map for the web page document of the textual contents therein. The secondary documents include one or more images with textual content embedded therein. An optical character recognition engine is used on the images to extract textual content for adding to the textual map for the webpage document.

In order to more particularly point out this feature of a hub processing unit of properly indexing a web document at a given address which includes secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation and where the secondary documents include one or more images with textual content embedded therein, the following language has been added to the independent claims, i.e., claims 1, 14, and 20 as follows:

retrieving a web document at an address, and extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser;

loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein;

analyzing and summarizing the in-memory webpage representation to produce a text map for the web page document of the textual contents therein; and

using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document.

No new matter has been added. Support for this amended language is found in the application as originally filed and particularly at FIGs. 3-8 and pages 9-13.

Rejection under 35 U.S.C. §103(a) in view of Sanu and Adar

As noted above, the Examiner rejected claims 1-3, 10, 14-16, and 20 under 35 U.S.C. §103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar. Specifically, Sanu taken alone and/or in view of Adar is silent on:

retrieving a web document at an address, and extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser;

loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein;

analyzing and summarizing the in-memory webpage representation

to produce a text map for the web page document of the textual contents therein; and

Sanu discloses a web crawler and worker thread which examines a set of properties and text within the web document. See Sanu at col. 7, lines 27-36. The web crawler, as taught by Sanu, is not working on an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser but rather the source of the web page content itself.

Further, Adar is completely silent on an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser. Accordingly, independent claims 1, 14, and 20 of the present invention, as amended, distinguish over both the Sanu and Adar references for at least this reason.

The Examiner goes on to combine Sanu with Adar.<sup>1</sup> The Examiner recites 35 U.S.C. §103. The Statute expressly requires that obviousness or non-obviousness be determined for the claimed subject matter "as a whole," and the key to proper determination of the differences between the prior art and the present invention is giving full recognition to the invention "as a whole." The Sanu reference taken alone or in view of Adar simply does not suggest, teach or disclose the patentably distinct limitation of:

an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser.

---

<sup>1</sup> Applicants make no statement whether such combination is even proper.

Continuing further, when there is no suggestion or teaching in the prior art for a hub processing unit for "extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser" the suggestion can not come from the Applicant's own specification. The Federal Circuit has repeatedly warned against using the Applicant's disclosure as a blueprint to reconstruct the claimed invention out of isolated teachings of the prior art. See MPEP §2143 and Grain Processing Corp. v. American Maize-Products, 840 F.2d 902, 907, 5 USPQ2d 1788 1792 (Fed. Cir. 1988) and In re Fitch, 972 F.2d 160, 12 USPQ2d 1780, 1783-84 (Fed. Cir. 1992).

For the foregoing reasons, independent claims 1, 14, and 20 as amended distinguish over Sanu taken alone or in view of Adar. Claims 2-3, 10, and 15-16 depend from claims 1 and 14 respectively. Since dependent claims contain all the limitations of the independent claims, claims 2-3, 10, and 15-16 distinguish over Sanu taken alone and/or in view of Adar, as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar and Aganovic

As noted above, the Examiner rejected claims 4-6, and 17-19 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Aganovic et al. (U.S. Patent Number 6,105,042). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar and/or Aganovic. As the Examiner correctly states on page 6 of the Office Action, Sanu and Adar are silent on optical character recognition and goes on to combine Sanu, Adar and Aganovic.<sup>2</sup> Careful reading of the optical character recognition as taught by Aganovic is for

---

<sup>2</sup> Applicants make no statement whether such combination is even proper.

document input during document storage. See Aganovic at col. 7, lines 1-5. In contrast, the optical recognition as recited in independent claims 1, 14 and 20 is for using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document wherein the "contents of the web document" is rendered as "an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser" and "loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein." Alternatively, the present invention uses optical scanning on "an intermediate dynamically constructed in-memory webpage representation of the web document" to "extract textual content" embedded in images from the stored web page output. Accordingly, independent claims 1, 14, and 20 distinguish over Sanu and/or in view of Adar and/or in view of Aganovic for at least this reason as well.

Moreover, the Federal Circuit has consistently held that when a §103 rejection is based upon a modification of a reference that destroys the intent, purpose or function of the invention disclosed in the reference, such a proposed modification is not proper and the *prima facie* case of obviousness can not be properly made. See *In re Gordon*, 733 F.2d 900, 221 USPQ 1125 (Fed. Cir. 1984). Here the intent, purpose and function of Aganovic is to use optical character recognition as a document input for storage into a database. In contrast, the intent and purpose of the present invention is to use "optical character recognition on the images to extract textual content for adding to the textual map for the webpage document." The present invention is directed to document output after retrieving the web document from an address and not for storage of the web page contents. Further, the optical character recognition is performed on "an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser." This combination, as suggested by the Examiner, destroys the intent and

purpose of Aganovic's use of optical character recognition on documents for storage. Accordingly, independent claims 1, 14, and 20 of the present invention are distinguishable over Sanu and/or in view of Adar and/or in view of Aganovic for at least this reason as well.

For the foregoing reasons, independent claims 1, 14, and 20 as amended distinguish over Sanu taken alone or in view of Adar and/or in view of Aganovic. Claims 4-6, and 17-19 depend from claims 1 and 14 respectively. Since dependent claims contain all the limitations of the independent claims, claims 4-6, and 17-19 distinguish over Sanu taken alone and/or in view of Adar and/or in view of Aganovic as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon and Meyerzon

As noted above, the Examiner rejected claim 7 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), in further view of Meyerzon et al. (U.S. Patent No. 6,199,081) and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar and/or Meyerzon ('081) and Meyerzon ('314). As the Examiner correctly states on page 9 of the Office Action, Sanu and Adar are silent on scheduling a URL for crawling and goes on to combine Sanu and Adar with Meyerzon ('081) and Meyerzon ('314).<sup>3</sup>

For the reasons set forth in the section entitled "Rejection under 35 U.S.C. §103(a) in view of Sanu and Adar" above, independent claim 1 distinguishes over Sanu taken alone and/or in view of Adar. Specifically, Sanu and Adar are both silent on:

an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is

---

<sup>3</sup> Applicants make no statement whether such combination is even proper.



formatted as if displayed for viewing on an end-user's web browser

In a similar manner, Meyerzon ('081) and Meyerzon ('314) disclose a web crawler and worker thread which examines a set of properties and text within the web document. See Meyerzon ('081) at col. 9 lines 60 through col. 10, lines 27 and Meyerzon ('314) at col. 9, line 28 through col. 10, lines 12. The web crawler, as taught by Meyerzon, is not working on an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser but rather the source of the web page content itself. Accordingly, independent claims 1, 14, and 20 of the present invention as amended distinguish over Sanu and/or Adar and/or in further view of for Meyerzon ('081) and Meyerzon ('314) for at least this reason.

For the foregoing reasons, independent claims 1, 14, and 20 as amended distinguish over Sanu and/or Adar and/or in further view of for Meyerzon ('081) and Meyerzon ('314). Claim 7 depends from claim 1. Since dependent claims contain all the limitations of the independent claims, claim 7 distinguishes over Sanu and/or Adar and/or in further view of for Meyerzon ('081) and Meyerzon ('314) as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon ('081)

As noted above, the Examiner rejected claims 8-9 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), in further view of Meyerzon et al. (U.S. Patent No. 6,199,081). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar and/or Meyerzon ('081).

For the reasons set forth in the section entitled "Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon and Meyerzon" above, independent claim 1

ARC9-2000-0046-US1

17

09/607,370

distinguishes over Sanu and/or in view of Adar, and/or in further view of Meyerzon ('081).

For the foregoing reasons, independent claim 1 as amended distinguishes over Sanu and/or in view of Adar, and/or in further view of Meyerzon ('081). Claims 8-9 depend from claim 1. Since dependent claims contain all the limitations of the independent claims, claims 8-9 distinguish over Sanu and/or Adar and/or in further view of for Meyerzon ('081) as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon ('314)

As noted above, the Examiner rejected claim 11 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), and in further view of Meyerzon et al. (U.S. Patent No. 6,638,314). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar and/or Meyerzon ('314).

For the reasons set forth in the section entitled "Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon and Meyerzon" above, independent claim 1 distinguishes over Sanu and/or in view of Adar, and/or in further view of Meyerzon ('314).

For the foregoing reasons, independent claim 1 as amended distinguishes over Sanu and/or in view of Adar, and/or in further view of Meyerzon ('314). Claim 11 depend from claim 1. Since dependent claims contain all the limitations of the independent claims, claim 11 distinguishes over Sanu and/or Adar and/or in further view of for Meyerzon ('314) as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

ARC9-2000-0046-US1

18

09/607,370

Rejection under 35 U.S.C. §103(a) in view of  
Sanu, Adar, Meyerzon ('314), Hughes and Aganovic

As noted above, the Examiner rejected claims 12 and 13 under 35 U.S.C. 103(a) as being unpatentable over Sanu et al. (U.S. Patent No. 6,145,003), in view of Adar et al. (U.S. Patent No. 6,493,702), in further view of Meyerzon et al. (U.S. Patent No. 6,199,314), in further view of Hughes et al. (U.S. 5,892,908), and in further view of Aganovic (U.S. Patent No. 6,105,042). Independent claims 1, 14 and 20 have been amended to distinguish over Sanu taken alone and/or in view of Adar, and/or in further view of Meyerzon ('314), and/or in further view of Hughes and/or in further view of Aganovic. As the Examiner correctly states on page 14 of the Office Action, Sanu, Adar, Meyerzon ('314), and Aganovic are silent on JPEG and TIF images and goes on to combine Sanu, Adar, Meyerzon ('314), and Aganovic with Hughes.<sup>4</sup>

For the reasons set forth in the sections entitled "Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar and Aganovic" and "Rejection under 35 U.S.C. §103(a) in view of Sanu, Adar with Meyerzon and Meyerzon" above, independent claim 1 distinguishes over Sanu taken alone and/or in view of Adar and/or in view of Aganovic and/or in view of Meyerzon ('314). Specifically, Sanu and Adar and Aganovic and Meyerzon are silent on:

an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser

and

optical character recognition on the images to extract textual content for adding to the textual map for the webpage document"

For the foregoing reasons, Independent claims 1, 14, and 20 as amended distinguish over Sanu taken alone and/or in view of Adar, and/or in further view of Meyerzon ('314),

---

<sup>4</sup> Applicants make no statement whether such combination is even proper.

and/or in further view of Hughes and/or in further view of Aganovic. Claims 12 - 13 depend from claim 1. Since dependent claims contain all the limitations of the independent claims, claim 12 - 13 distinguish over Sanu taken alone and/or in view of Adar, and/or in further view of Meyerzon ('314), and/or in further view of Hughes and/or in further view of Aganovic as well, and the Examiner's rejection should be withdrawn, which withdrawal is respectfully requested.

### CONCLUSIONS

The remaining cited references have been reviewed and are not believed to effect the patentability of the claims as previously amended.

In light of the Office Action, Applicants believe these amendments serve a useful clarification purpose, and are desirable for clarification purposes, independent of patentability. Accordingly, Applicants respectfully submit that the claim amendments do not limit the range of any permissible equivalents.

Applicants acknowledge the continuing duty of candor and good faith to the disclosure of information known to be material to the examination of this application. In accordance with 37 CFR §§ 1.56, all such information is dutifully made of record. The foreseeable equivalents of any territory surrendered by amendment is limited to the territory taught by the information of record. No other territory afforded by the doctrine of equivalents is knowingly surrendered and everything else is unforeseeable at the time of this amendment by the Applicants and their attorneys.

Applicants respectfully submit that all of the grounds for rejection stated in the Examiner's Office Action have been overcome, and that all claims in the application are allowable. No new matter has been added. It is believed that the application is now in condition for allowance, which allowance is respectfully requested.

ARC9-2000-0046-US1

20

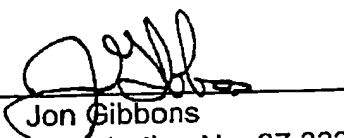
09/607,370

**PLEASE**, if for any reason the Examiner finds the application other than in condition for allowance, the Examiner is invited to call either of the undersigned attorneys at (561) 989-9811 should the Examiner believe a telephone interview would advance the prosecution of the application.

Respectfully submitted,

Date: June 23, 2004

By: \_\_\_\_\_

  
Jon Gibbons  
Registration No. 37,333  
Attorney for Applicants

FLEIT, KAIN, GIBBONS,  
GUTMAN, BONGINI & BIANCO P.L.  
One Boca Commerce Center, Suite 111  
551 Northwest 77th Street  
Boca Raton, FL 33487  
Tel. (561) 989-9811  
Fax (561) 989-9812

Please Direct All Future Correspondence to Customer Number **23334**

ARC9-2000-0046-US1

21

09/607,370